The Canadian Journal of Statistics
Vol. 17, No. 2, 1989, Pages 235–239
La Revue Canadienne de Statistique

Consistent deconvolution in density estimation*

LUC DEVROYE

McGill University

Key words and phrases: Density estimation, deconvolution, random noise, signal detection, convergence, consistency, characteristic function.
AMS 1985 subject classifications: 62G05

ABSTRACT

Suppose we have n observations from $X = Y + Z$, where Z is a noise component with known distribution, and Y has an unknown density f. When the characteristic function of Z is nonzero almost everywhere, we show that it is possible to construct a density estimate f_n such that for all f, $\lim_{n\to\infty} \mathcal{E} \int |f_n - f| = 0$.

RÉSUMÉ

Supposons que l'on dispose de n observations de la variable $X = Y + Z$, où Z est un bruit aléatoire de loi connue et Y possède une densité, f, inconnue. Il est démontré que si la fonction caractéristique de Z est non nulle presque partout, alors on peut construire une estimation, f_n, telle que pour toute densité f, $\lim_{n\to\infty} \mathcal{E} \int |f_n - f| = 0$.

1. INTRODUCTION AND MAIN RESULT

The deconvolution problem is of interest in various areas of engineering statistics, such as the problem of the extraction of a signal when one observes signal plus noise. It is well known how to consistently estimate a density f on the basis of an independently and identically distributed sample $Y_1, \ldots, Y_n$. But can we still do so if we do not observe the Y_i's, but rather the noise-perturbed values $X_1, \ldots, X_n$, where $X_i \equiv Y_i + Z_i$, and the Z_i's form an independently and identically distributed "noise" sequence with a known noise distribution? Around 1985, G. Tusnády (personal communication to L. Györfi) asked if one could find an estimate f_n such that for all f, under normal (0,1) noise, $\mathcal{E} \int |f_n - f| \to 0$. In this note, we point out that we can consistently estimate the density whenever the noise has a characteristic function ζ with the property that $\zeta(t) \neq 0$ for almost all t. In particular, the noise need not even be absolutely continuous. The estimate constructed in the proof is based on the kernel estimate, but differs from the deconvolution kernel estimates recently proposed in the literature. Deconvolution kernel estimates were proposed as early as 1977 rece by Wise, Traganitis, and Thomas for Poisson noise. For normal noise, see the t study by Carroll and Hall (1988). Devroye and Wise (1977) looked at the L_∞ behaviour of estimates of distributions on the integers when the noise

*This research was supported by a Natural Sciences and Engineering Research Council Grant No. A 3456.

is also integer-valued. Stefanski and Carroll (1987) and Liu and Taylor (1987) studied deconvolution kernel estimates in depth [see also Stefanski (1988) and Fan (1988)].

THEOREM. *Let $\zeta(t) \neq 0$ for almost all t. Then it is possible to construct an estimate f_n such that for all f,*

$$\lim_{n\to\infty} \mathcal{E} \int |f_n - f| = 0.$$

REMARK. If ζ violates the conditions of the theorem, then we can find and infinite family of densities f such that for any f in this family, $X_1 = Y_1 + Z_1$ has the same density. Just note that by the continuity of ζ, there exists an interval $[a,b]$ on the positive half line such that $\zeta \equiv 0$ on this interval and on its symmetric counterpart. Now choose a density f with a nonzero Pólya characteristic function ψ (i.e., ψ is real and convex on the positive half line). Clearly, the characteristic function of X_1 is $\psi\zeta$, so that changing ψ on $[a,b]$ does not affect the distribution of X_1. For $\psi(t) = e^{-|t|}$, for example, this can be done in an infinite number of ways, as long as we keep the modified characteristic function symmetric and convex [fit a linear piece between $\psi(a)$ and $\psi(b)$, and consider mixtures to obtain the infinite family]. Thus, consistent estimation in general is impossible in such cases.

2. DEFINITION OF THE ESTIMATOR

Consider the function

$$\rho(u) \triangleq \inf_{|t|\leq u} |\zeta(t)|,$$

and the sets $A_u = \{t : |\zeta(t)| < u\}$. Let λ denote Lebesgue measure. The estimator we propose requires a *tail parameter* $T > 0$, a *smoothing parameter* $h > 0$, a *noise-control parameter* $r \geq 0$, and a *kernel* K, which is an integrable function with $\int K = 1$ whose Fourier transform $\psi(t) \triangleq \int e^{itx} K(x)\, dx$ satisfies

$$\sup_t |\psi(t)| \leq C < \infty, \tag{1}$$

$$\psi(t) \equiv 0 \quad \text{for} \quad t \notin [-c,c], \quad c < \infty. \tag{2}$$

A possible kernel is the de la Vallée Poussin kernel $K(x) = (1/2)\pi)[(\sin \frac{1}{2}x)/\frac{1}{2}x]^2$, which has Fourier transform $\psi(t) = (1-|t|)_+$, $c = 1$, and $C = 1$. All parameters are allowed to vary with n. However, they must satisfy the following growth conditions:

$$\lim_{n\to\infty} T = \infty, \qquad \lim_{n\to\infty} h = 0, \tag{3}$$

$$\lim_{n\to\infty} \frac{T}{nh \max^2(\rho(c/h), r)} = 0, \tag{4}$$

and

$$\lim_{n\to\infty} T\lambda\left(A_r \cap \left[-\frac{c}{h}, \frac{c}{h}\right]\right) = 0. \tag{5}$$

Our estimator f_n generalizes the kernel estimate of Parzen (1962) and Rosenblatt (1956):

$$f_n(x) = \begin{cases} 0 \\ \dfrac{1}{2\pi} \operatorname{Re}\left\{\displaystyle\int_{R-A_r} e^{-itx}\psi(th)\zeta^{-1}(t)\phi_n(t)\, dt\right\} & \text{if} \quad |x| < T, \end{cases}$$

where $\phi_n(t) = (1/n)\sum_{j=1}^{n} e^{itX_j}$ is the empirical characteristic function for $X_1, \ldots, X_n$. The two differences with estimators found in Stefanski and Carroll (1987) or Liu and Taylor (1987) are the truncation device T and the inversion by integrating over $R - A_r$ only. The motivation is quite simple: for small t, ϕ_n is close to the characteristic function of X_1, i.e. $\phi\zeta$. Substituting this in the definition for f_n, we see that for $|x| \le T, f_n(x)$ is close to $f * K_h(x) = (1/2\pi)\int e^{-itx}\psi(th)\phi(t)\,dt$, where $*$ is the convolution operator, and $K_h(z) = (1/h)K(z/h)$. For small h, this is close to f in several respects; for example $\lim_{h\downarrow 0}\int |f - f * K_h| = 0$ for all f (Wheeden and Zygmund, 1977).

3. PROOF OF THE THEOREM

First, we show that it is always possible to find T, h, r such that (3–5) are satisfied. For integer k, we set $h = 1/k$ and define $r = \sup\{u : \lambda(A_u \cap [-c/h, c/h]) < 1/k\}$. Such an r exists since for every constant $\nu > 0$, $\lambda(A_u \cap [-\nu, \nu]) \longrightarrow 0$ by our assumption concerning ζ. Then set $n_k = \lceil k/(hr^2)\rceil$. For $n \in [n_k, n_{k+1})$, define h and r as indicated. Note that $n_k \uparrow \infty$, $h \longrightarrow 0$ and $\lambda(A_r \cap [-c/h, c/h]) \longrightarrow 0$ as $k \uparrow \infty$. Furthermore, $nhr^2 \longrightarrow \infty$ as $n \longrightarrow \infty$. Define $T^2 = \min(nhr^2, 1/\lambda(A_r \cap [-c/h, c/h]))$, so that $T \longrightarrow \infty$, and (4) and (5) are satisfied as well.

Let us define the function

$$q_n(x) = \frac{1}{2\pi}\mathrm{Re}\int_{R-A_r} e^{-itx}\psi(th)\phi(t)\,dt.$$

We note that

$$\begin{aligned}
\int |f_n - f| &\le \int_{|x|\le T} |f_n - f| + \int_{|x|>T} f \\
&\le \int_{|x|\le T} |f_n - q_n| + \int_{|x|\le T} |f - q_n| + \int_{|x|>T} f \\
&\le \sqrt{2T}\left(\int_{|x|\le T} (f_n - q_n)^2\right)^{1/2} + 2T\sup_x |q_n - f^*K_h| + \int |f - f^*K_h| + \int_{|x|>T} f \\
&\stackrel{\Delta}{=} \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.
\end{aligned}$$

Clearly, IV $\longrightarrow 0$ as $T \longrightarrow \infty$, and III $\longrightarrow 0$ as $h \longrightarrow 0$. Furthermore,

$$\mathrm{II} \le \frac{T}{\pi}\int_{A_r} |\psi(th)|\,|\phi(t)|\,dt \le \frac{TC\lambda(A_r \cap [-c/h, c/h])}{\pi} \longrightarrow 0.$$

Finally, let us introduce the notation

$$\begin{aligned}
L_{nj}(x) &= \frac{1}{2\pi}\int_{R-A_r} e^{it(X_j - x)}\psi(th)\zeta^{-1}(t)\,dt, \\
K_{nj}(x) &= \mathrm{Re}\, L_{nj}(x).
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{E}^2(I) &\le 2T\int_{|x|\le T} \mathcal{E}(f_n - q_n)^2 \le \frac{2T}{n}\mathcal{E}\int K_{n1}^2(x)\,dx \le \frac{2T}{n}\mathcal{E}\int L_{n1}^2(x)\,dx \\
&= \frac{T}{\pi n}\int_{R-A_r} |\psi(th)|^2|\zeta^{-1}(t)|^2\,dt \qquad \text{(Parseval's identity)} \\
&\le \frac{2cTC^2}{\pi nh\,\max^2(\rho(c/h), r)}\,.
\end{aligned}$$

4. REMARKS, EXTENSIONS, SPECIAL CASES

(a) *No noise.* In the absence of noise, we have $\zeta \equiv 1$ and $\rho \equiv 1$, so that (5) is satisfied for all constant $r < 1$, and (4) holds when $nh/T \longrightarrow \infty$. In fact, (3) and (4) together hold if we start with any standard sequence h ($h \longrightarrow 0$ and $nh \longrightarrow \infty$ are the "standard" conditions on h), and define $T = \sqrt{nh}$.

(b) *Weaker conditions.* Observe that (5) can be replaced by the condition

$$\lim_{n\to\infty} T \int_{|t|<c/h, T\in A_r} |\phi(t)|\, dt = 0. \tag{6}$$

Unfortunately, this condition involves the unknown ϕ, and is difficult to ensure beforehand. Note however that (5) implies (6).

(c) ϕ *is absolutely integrable.* Let h be any standard sequence (i.e. $h \longrightarrow 0$ and $nh \longrightarrow \infty$). Then a little work shows that we can always find r and T to satisfy (3)–(4) and (6). To see this, in the proof of the theorem, take $n_k = k^4$, and $r = h = 1/k$ on $[n_k, n_{k+1}]$. Thus, for absolutely integrable ϕ, *any* consistent kernel estimate with a kernel as used in the theorem which is consistent in the noiseless case can be modified without altering h or K by the T and r devices to yield a consistent estimate in the noisy case.

(d) ϕ *is symmetric and strictly monotonically decreasing on the positive half line.* Here we can take $r = \phi^{\text{inv}}(c/h)$, so that (5) automatically holds. The existence of a sequence T satisfying (3)–(4) now follows from the conditions $h \longrightarrow 0$ and either $nh\phi^{\text{inv}}(c/h) \longrightarrow \infty$ or $nh\rho(c/h) \longrightarrow \infty$.

(e) $\zeta(t) \neq 0$ *for all* t. In this case, (3)–(5) can be replaced by (3) and the condition $T/nh\rho(c/h) \longrightarrow 0$.

(f) *Normal noise.* The condition (4) is satisfied when $h\sqrt{\log n} \longrightarrow \infty$ and $T = O(n^{1-\epsilon})$ for some $\epsilon > 0$.

(g) *Choice of the kernel.* It is well known that the best kernels are those whose characteristic function is the flattest near the origin [see Bullock-Davis (1975, 1977) or Devroye (1987)]. It seems that in deconvolution density estimation, the best rates are also obtained when ψ is flat near the origin (e.g., $\psi \equiv 1$ in an open neighborhood of the origin), which should be added to the condition that ψ has compact support. One could for example consider the trapezoidal characteristic functions suggested in Devroye and Györfi (1985).

(h) *Strong convergence.* L. Györfi has pointed out to me that the techniques of Devroye (1988) can be used to show that the estimator presented above satisfies $\int |f_n - f| \longrightarrow 0$ almost surely when (3) and (5) hold, ζ satisfies the condition of the theorem, and (4) is replaced by

$$\lim_{n\to\infty} \frac{T\sqrt{\log n}}{nh \max^2(\rho(c/h), r)} = 0. \tag{7}$$

ACKNOWLEDGEMENT

I would like to thank Laci Györfi for some helpful suggestions and for his profound insight.

REFERENCES

Carroll, R.J., and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* 83(404), 1184–1186.

Bullock-Davis, K. (1975). Mean square error properties of density estimates. *Ann. Statist.*, 5, 1025–1030.

Bullock-Davis, K. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.*, 5, 530–535.

Devroye, L. (1987). *A Course in Density Estimation.* Birkhäuser, Boston.

Devroye, L. (1988). The kernel estimate is relatively stable. *Probab. Theory Related Fields*, 77, 521–536.

Devroye, L., and Györfi, L. (1985). *Nonparametric Density Estimation: The L1 View.* Wiley, New York.

Devroye, L., and Wise, G.L. (1977). On the estimation of discrete probability densities from noisy measurements. *Proc. 15th Annual Allerton Conference on Communication, Control and Computing*, 211–220.

Fan, J. (1988a). On the optimal rates of convergence for nonparametric deconvolution problem. Manuscript, Department of Statistics, University of California at Berkeley, 1988.

Fan, J. Optimal global rates of convergence for nonparametric deconvolution problem. Manuscript, Department of Statistics, University of California at Berkeley, 1988.

Liu, M.C., and Taylor, R.L. A consistent nonparametric density estimator for the deconvolution problem. Technical Report 73, Department of Statistics, University of Georgia, 1987.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.*, 33, 1065–1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832–837.

Stefanski, L.A. (1988). Rates of convergence of some estimators in a class of deconvolution problems. Institute of Statistics Mimeograph Series 1935, North Carolina State University, Raleigh, N.C.

Stefanski, L.A., and Carroll, R.J. (1987). Deconvoluting kernel density estimators. University of North Carolina Mimeo Series, 1623.

Wheeden, R.L., and Zygmund, A. (1977). *Measure and Integral.* Mercel Dekker, New York.

Wise, G.; Traganitis, A.P., and Thomas, J.B. (1977). The estimation of a probability density function from measurements corrupted by Poisson noise. *IEEE Trans. Inform. Theory*, IT-23, 764–766.

Received 15 September 1988
Revised 9 March 1989
Accepted 15 March 1989

School of Computer Science
McGill University
Montréal, Québec H3A 2K6